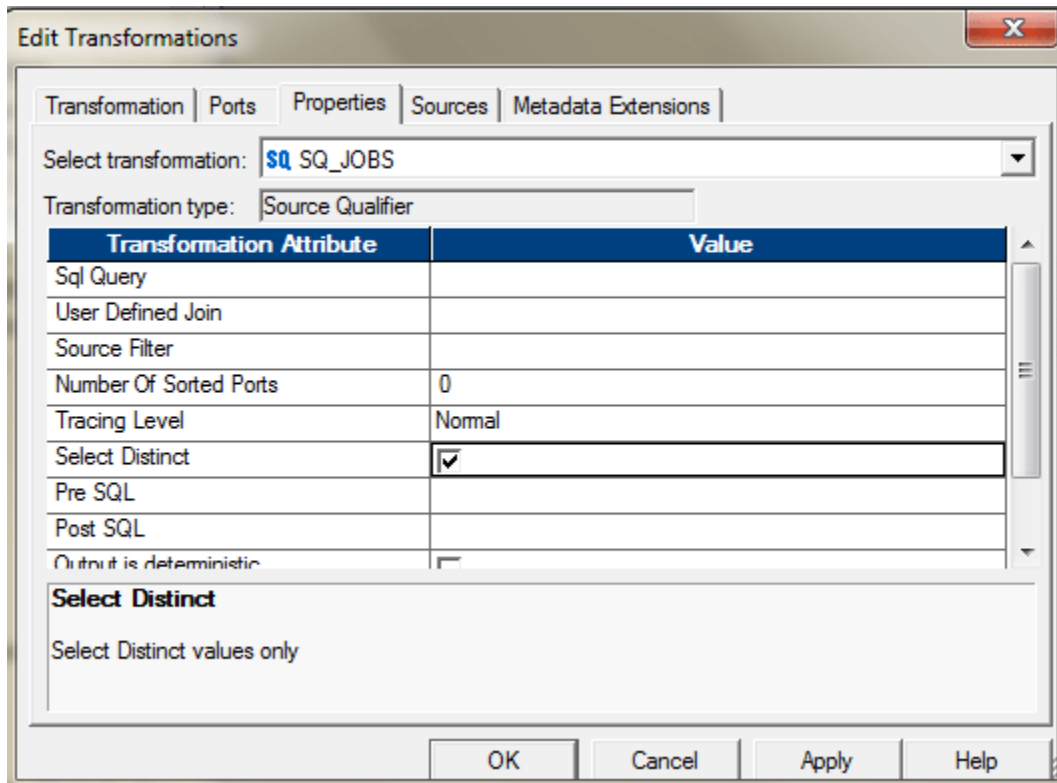


How do you remove Duplicate records in Informatica? And how many ways are there to do it?

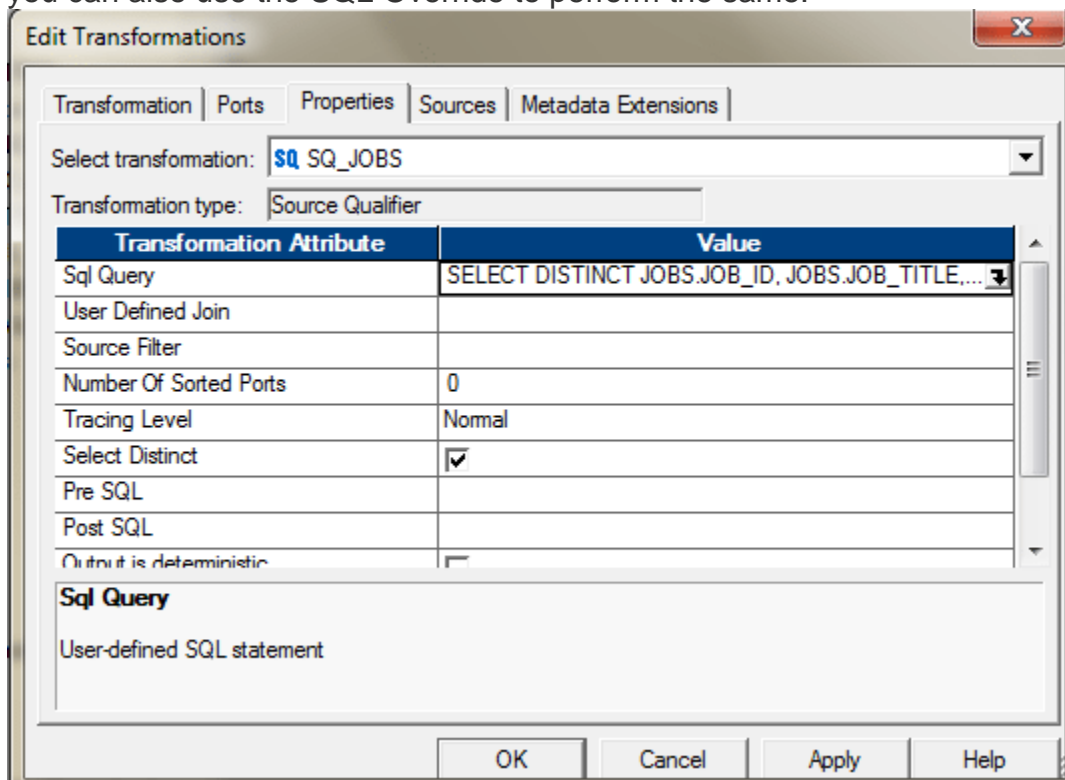
There are several ways to remove duplicates.

- i. If the source is DBMS, you can use the property in Source Qualifier to select the distinct records.



Or

you can also use the SQL Override to perform the same.



- ii. You can use, Aggregator and select all the ports as key to get the distinct values. After you pass all the required ports to the Aggregator, select all those ports , those you need to select for de-duplication. If you want to find the duplicates based on the entire columns, select all the ports as group by key.

Edit Transformations

Transformation Ports Properties Metadata Extensions

Select transformation: **agg_deduplication**

Transformation type: **Aggregator**

	Port Name	Datatype	Prec	S...	I	O	V	Expression	Grou...
1	EMPLOYEE_ID	decimal	6	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	EMPLOYEE_ID	<input checked="" type="checkbox"/>
2	FIRST_NAME	nstring	20	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	FIRST_NAME	<input checked="" type="checkbox"/>
3	LAST_NAME	nstring	25	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	LAST_NAME	<input checked="" type="checkbox"/>
4	EMAIL	nstring	25	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	EMAIL	<input checked="" type="checkbox"/>
5	PHONE_NUMBER	nstring	20	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	PHONE_NUMBER	<input checked="" type="checkbox"/>
6	HIRE_DATE	date/time	29	9	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	HIRE_DATE	<input checked="" type="checkbox"/>
7	JOB_ID	nstring	10	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	JOB_ID	<input checked="" type="checkbox"/>
8	SALARY	decimal	8	2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SALARY	<input checked="" type="checkbox"/>

Default value:

Description:

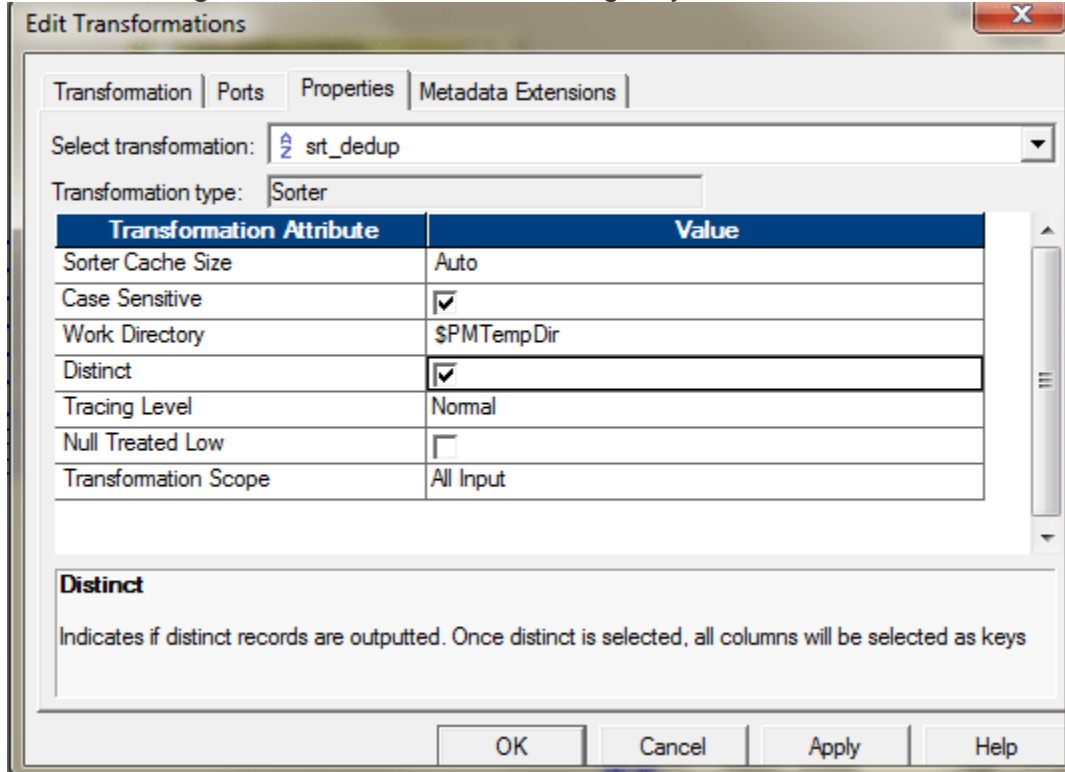
OK Cancel Apply Help

The

Mapping will look like this.

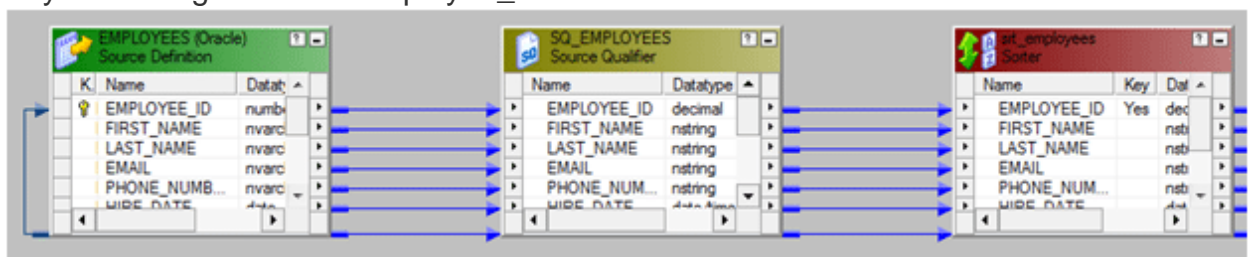


- iii. You can use Sorter and use the Sort Distinct Property to get the distinct values. Configure the sorter in the following way to enable this.

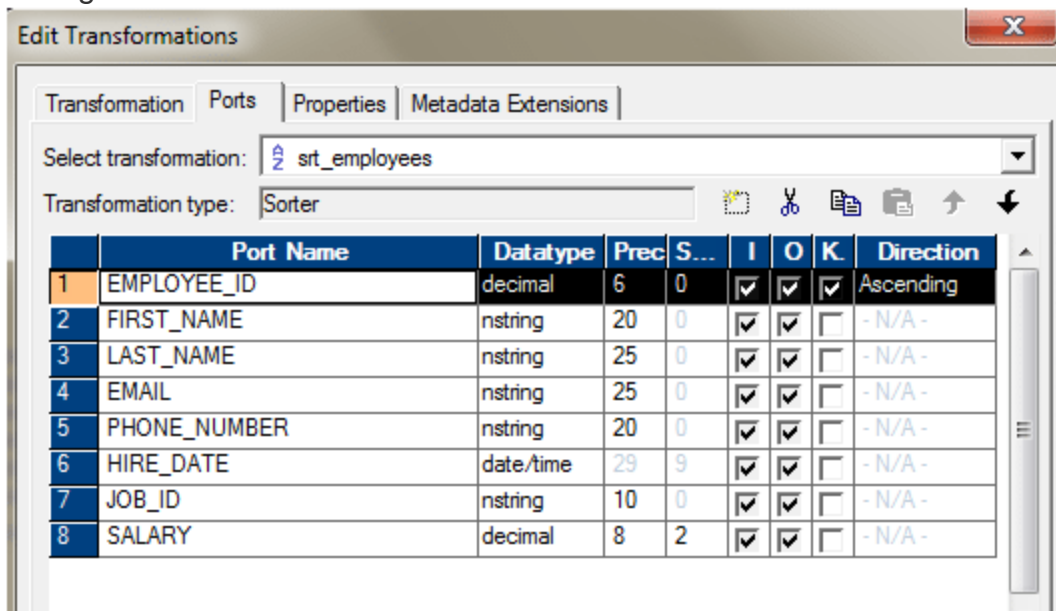


- iv. You can use, Expression and Filter transformation, to identify and remove duplicate if your data is sorted. If your data is not sorted, then, you may first use a sorter to sort the data and then apply this logic:

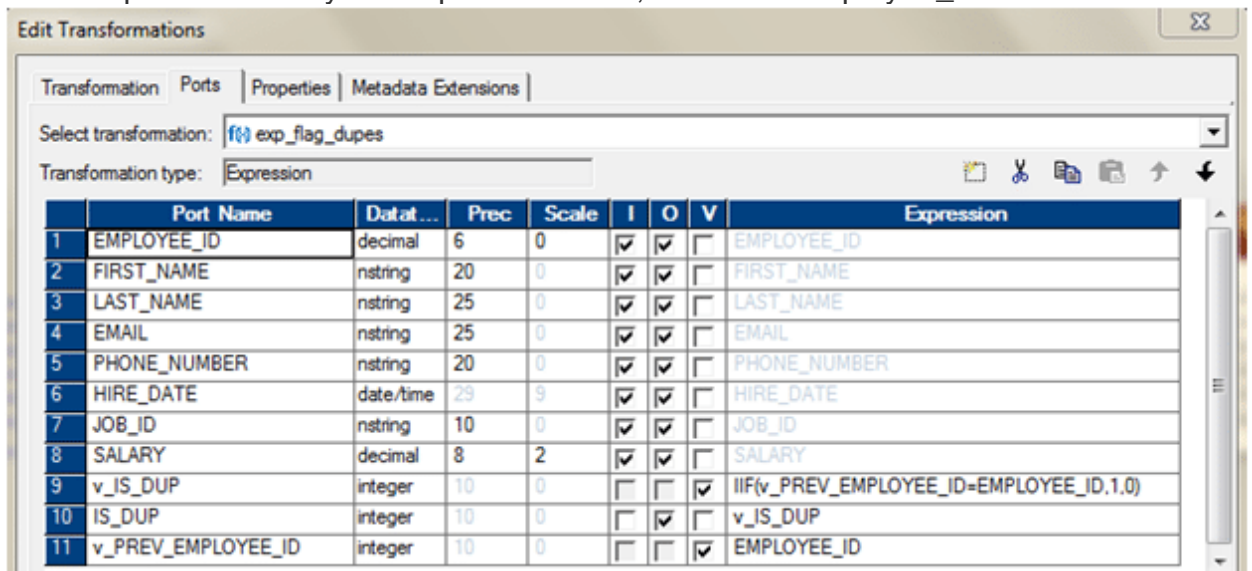
- Bring the source into the Mapping designer.
- Let's assume the data is not sorted. We are using a sorter to sort the data. The Key for sorting would be Employee_ID.



Configure the Sorter as mentioned below.

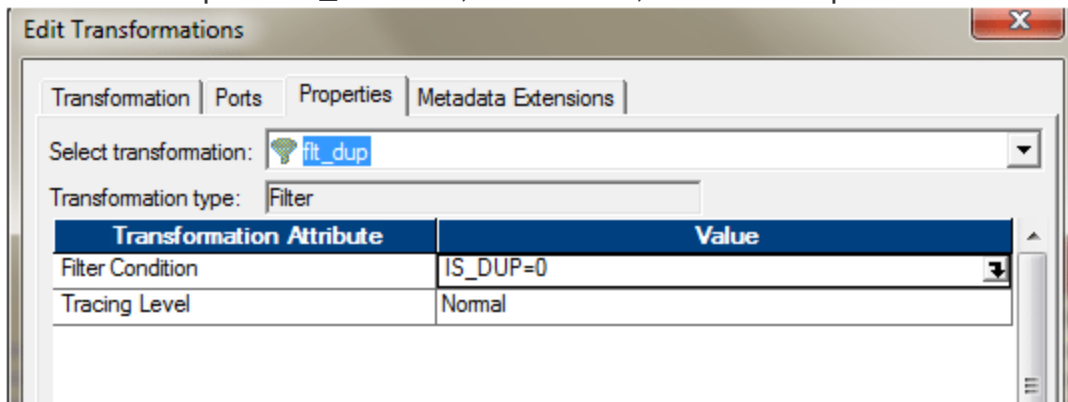


- Use one expression transformation to flag the duplicates. We will use the variable ports to identify the duplicate entries, based on Employee_ID.



- Use a filter transformation, only to pass IS_DUP = 0. As from the previous expression transformation, we will have IS_DUP =0 attached to only records,

which are unique. If $IS_DUP > 0$, that means, those are duplicate entries.



- Add the ports to the target. The entire mapping should look like this.



v. When you change the property of the Lookup transformation to use the Dynamic Cache, a new port is added to the transformation. NewLookupRow.

The Dynamic Cache can update the cache, as and when it is reading the data.

If the source has duplicate records, you can also use Dynamic Lookup cache and then router to select only the distinct one.